What Decides Resale Value for Vehicles in the UK?

Lab 2 Report - Team Herkimer: Rick Chen, Chase Madson, Maria Manna, Jash Sompalli

Introduction

In the past two years, new vehicle shortages have created higher demand for pre-owned vehicles across the globe. New vehicle sales have been impacted by mismatched inventory and manufacturing constraints due to semiconductor shortages and supply chain failures, and with the substantial increase in vehicle costs, consumers have had little choice but to embrace the pre-owned car market. This creates an opportunity for Acme, Inc., the largest used car dealer in the United Kingdom, to capitalize on the increased demand for their product. Analyzing used vehicle listings will shape the future decision making in asset acquisition, avoiding vehicles unfairly priced above estimated market value.

This study examines the characteristics and history of pre-owned vehicles to determine impact on market value with the intent of guiding companies to acquire appropriately priced vehicles with the best chances of resale. The data in our analysis is comprised of used car listings as distinct observations, for each of which we have information on the market value price and our causal variables of interest - vehicle age and mileage. We also consider the fixed effects of characteristics like vehicle type, fuel economy, and transmission.

Data and Methodology

The dataset utilized in our analysis, **100,000 UK Used Car Data Set**, contains observations of used car listings scraped from the web by Kaggle user ADITYA. Following data cleansing, each of the remaining 99,010 rows in the compiled dataset represent an active used car listing on the internet when the data was scraped in 2020. Data exploration was performed on a 30% sample of the data and the remaining observations were utilized to generate our report statistics. During our initial data exploration, we noted a few oddities. There were 247 observations that listed the fuel type of the vehicle as "other." This was likely due to a flaw in the collection of data as none of the vehicles ran via alternative fuels such as biodiesel or hydrogen fuel cell; all should have been classified as either petrol, diesel, electric, or hybrid. Due to the small number of observations affected, they were removed from our dataset. Additionally, there were 9 observations that listed the transmission type as "other." We believe this to be a mistype as these observations could have been classified as manual, automatic, or semi-automatic, therefore they were also removed from the dataset.

Lastly, we removed <1% of the total observations due to having reported unreasonable fuel economy (<25 or >90 Imperial MPG). In mapping the causal theory of our model, we initially chose to include engine size. However, there were many inaccuracies in this category, such as reporting an engine size of zero for a hybrid vehicle or an engine size greater than zero for an electric vehicle. Due to this, we removed engine size from our model. Therefore, age, mileage, vehicle type, transmission type, fuel type, fuel economy, and manufacturer country were identified as variables with the potential to impact market price (Figure 1).

After evaluating these factors, we noted that vehicle mileage was an outcome variable of vehicle age. Therefore, in order to still capture the information on level of vehicle use, we strategically created a variable for annual mileage that broke the link between age and overall mileage. We transformed mileage by dividing mileage by vehicle age, and then binned the outcomes into four brackets of annual mileage, using the insurance domain standards of low (<5,000), medium, high (>10,000), and a fourth category for vehicles that are less than one year old and their annual mileage is yet to be determined (TBD). This is depicted in our final causal map (Figure 2).

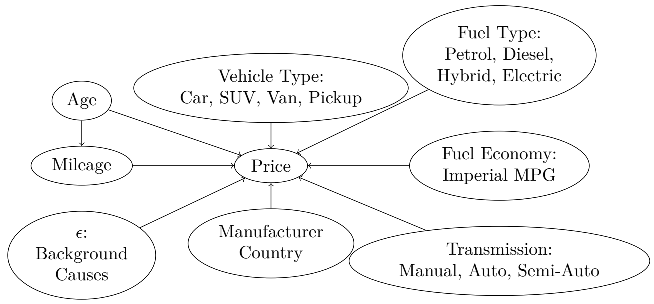

Figure 1: Basic Causal Map

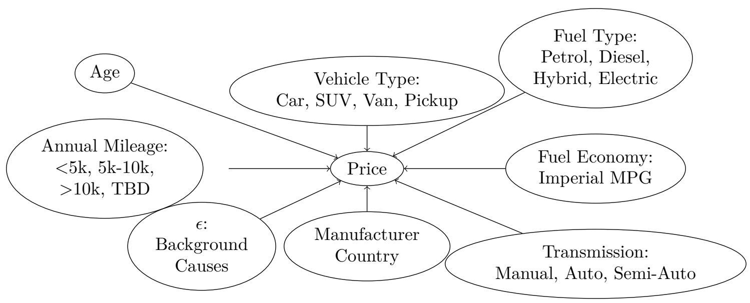

Figure 2: Causal Map with Annual Mileage Brackets

Mathematically, the equation for our final regression model takes the form:

$$Price = \beta_0 + \beta_1 \times Age + (\beta_2 \times Annual\ Mileage_{<5k} + \beta_3 \times Annual\ Mileage_{>10k} + \beta_4 \times Annual\ Mileage_{TBD}) \\ + \Sigma_{i=5}^{9}(\beta_i \times Baseline\ Attribute\ FixedEffects_i) + \epsilon$$

where $\beta_1 < 0$ represents the average decrease in market price for every year of a vehicle's age, $\beta_2 > 0$ represents the average increase in price for a car that has *relatively low mileage* for its age, $\beta_3 < 0$ represents the average decrease in price for a car that has *relatively high mileage* for its age, and $\beta_4 > 0$ represents the average increase in price for a car that's less than 1 year old and we cannot yet calculate its annual mileage. These variables affect the resale price of a vehicle and reflect its usage since it was first sold. β_5 through β_9 are a set of fixed effects corresponding to baseline attributes known about the car at the time of manufacturing (e.g., transmission type, fuel type) and their effects on price that we wish to control for. ϵ represents the unmeasured factors that have a causal effect on price.

Results

The results of our regression analysis are captured in Table 1. Three regression models were tested, and across all models the coefficients were all highly statistically significant. In Figure 3 we plot vehicle price against its age, where heavily driven cars appearing in blue, lightly driven cars in red, and moderates in green. Clearly we see price diminish with age, with heavily driven cars depreciating the fastest.

As we progressed through each of the three models, we replicated the process of a consumer searching for a used car, starting with the most important considerations (age, annual mileage) and adding features with each subsequent model. Moving from our base model to Model 2, we added in a variable for fuel efficiency, measured in Imperial MPG. We progressively added in fixed effect variables for vehicle type and fuel efficiency (Models 2, 3), and then variables for transmission type, fuel type, and origin country (Model 3). As shown

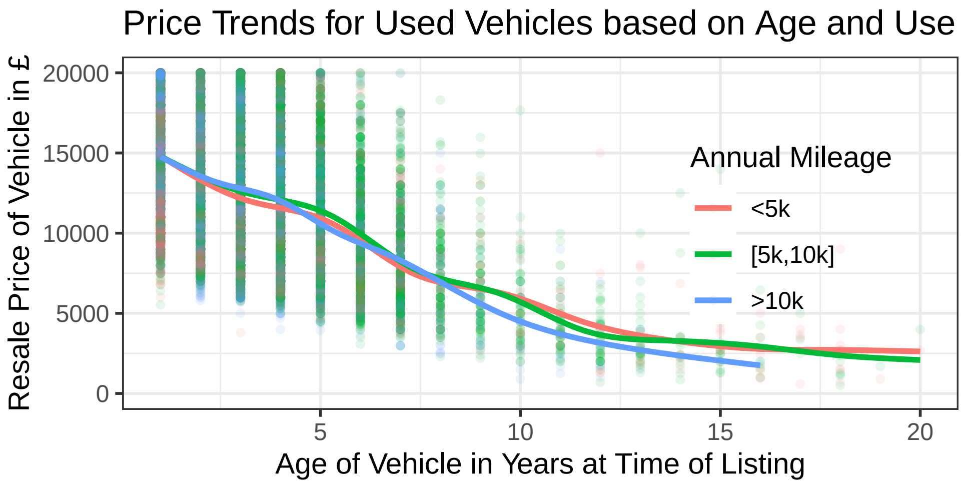

Figure 3: Used Car Prices Diminish with Age, Especially for Highly Driven Cars

in the table below, all three of the models pass the F-statistic test (p-value < 0.01) and each coefficient is statistically significant. Taking into account each of the fixed effects variables, we end up with an R^2 of 0.7367. In parentheses under each term is a 95% confidence interval calculated using robust standard errors.

Table 1: Regressing Vehicle Price Against Age, Annual Mileage Bracket, and Other Vehicle Attributes

	Resale Price of Vehicle		
	(1)	(2)	(3)
Age of Vehicle in Years	$-2{,}090.47^{***}$	$-1{,}791.18^{***}$	$-1{,}727.78^{***}$
	($-2{,}140.84$, $-2{,}040.10$)	($-1{,}830.66$, $-1{,}751.69$)	($-1{,}762.34$, $-1{,}693.22$)
Annual Mileage <5k	$1{,}031.15^{***}$	504.38^{***}	927.21^{***}
	(866.39, 1,195.90)	(361.41, 647.35)	(830.37, 1,024.06)
Annual Mileage >10k	$-1{,}431.17^{***}$	-833.46^{***}	$-1{,}550.84^{***}$
	($-1{,}562.43$, $-1{,}299.91$)	(-955.16, -711.75)	($-1{,}631.90$, $-1{,}469.79$)
Mileage TBD: 2020 Model	$5{,}338.29^{***}$	$3{,}790.67^{***}$	$1{,}881.53^{***}$
	(4,873.72, 5,802.87)	(3,387.09, 4,194.24)	(1,581.94, 2,181.12)
Constant	$22{,}548.39^{***}$	$36{,}352.34^{***}$	$39{,}939.07^{***}$
	(22,346.32, 22,750.47)	(35,913.56, 36,791.11)	(39,536.59, 40,341.55)
Vehicle Type FE	-	✓	✓
Fuel Economy FE	-	✓	✓
Transmission FE	-	-	✓
Fuel Type FE	-	-	✓
Manufacturing Country FE	-	-	✓
Observations	68,495	68,495	68,495
R^2	0.28	0.41	0.74
F Statistic	$6{,}645.66^{***}$ (df = 4; 68490)	$6{,}068.34^{***}$ (df = 8; 68486)	$11{,}267.43^{***}$ (df = 17; 68477)
Note:			*p<0.1; **p<0.05; ***p<0.01

Ultimately, in the current market and given the shortages we are facing, we find that a vehicle generally loses £1,728 in value for each year since manufacturing. If the annual mileage of the vehicle is high, it loses £1,551 in value, but it gains £927 in value if the annual mileage is low. If the vehicle is less than one year old, and therefore does not have an annual mileage (in terms of our calculation), we find the value of the vehicle to increase by £1,882. Knowing this information can inform Acme's decisions to purchase assets by understanding future profitability based on current price of acquisition - showing the practical significance of our model. To maximize profit, we should implement this data-based approach in selecting cars.

Limitations

This model has a few limitations that should be addressed. First, in order to run this model our units of observation must be independent and identically distributed (**IID**). It is possible, however, that *geographical clustering* exists within the data. Our model accounts for the manufacturer's country but data was only collected on manufacturers from six countries, with the majority of vehicles originating from Germany. Due to this, future generalizability of the model could be impacted. Also impacting future generalizability is the origin of the data itself. As the model was based on sales in the UK, if a subsidiary from a different country utilized the same model there may be important factors omitted in the analysis, such as increased maintenance costs of European vehicles due to a lack of readily available parts outside of Europe. Additionally, the model doesn't take into account how long it took for the vehicles to sell or the cost of housing an asset that isn't selling for a period of time. This would cut into profits and frequency of stock rotation, and could be a major factor in choosing particular vehicles based on the company's current needs. We observe right skew in our metric variables, but we believe their variance is finite and thus, noting the absence of perfect colinearity, we affirm our **unique BLP assumption**.

Reverse causality is not a concern in this model. It is not plausible for used vehicle's resale price to determine how it was manufactured years before. There may be a minority of cars with smaller annual mileage due to owners wanting to preserve the price, but we feel this would not be strong enough to negatively impact the explanatory power of our model. In our initial causal model, we considered including age and mileage simultaneously until it became clear that mileage was significantly influenced by the age of the vehicle. To avoid having an **outcome on the right-hand side**, we created an alternative variable for annual mileage of the vehicle, which eliminates the connection between age and mileage. For this research, we were limited to the data we collected which naturally omitted a lot of information about each used vehicle listed. Theoretically, this is encapsulated in the epsilon of our causal map, however it is important to note that some variables that are not included in the analysis may have a strong impact on the price of the vehicle. The most notable of these omitted variables would be the *number of accidents in each vehicle's history*, which would be correlated negatively with price (and potentially positively with its age). As a result, this is likely producing **omitted variable bias** in our model tilting our negative coefficient for age away from zero.

Conclusion

This regression study empowers Acme, Inc. to confidently select an appropriate offer price based on a given vehicle's usage attributes, considering current demand for certain types of vehicles. This in turn optimizes the acquisition of assets for the company by creating the best opportunity for higher profit margins. Once a potential buyer determines selection criteria for a vehicle (e.g., type of vehicle, fuel economy), this model can be used to assess a fair offer price for the vehicle. Acme, or any potential buyer leveraging this model, should first consider the depreciating £1,728 that each additional year of age has on a used car. Then, consider that a car with relatively small mileage for its age (<5k per year) is worth £927 more on average, while a car with high annual mileage (>10k per year) is worth £1,551 less on average. In the future, this could be utilized by other subsidiary online pre-owned car platforms, auto insurance companies, and individual consumers who want to purchase the most fairly-priced vehicles on the current market. As the composition of vehicles changes, whether that be a migration to hybrid and electric vehicles or an increased demand for utility vehicles, this model should be able to adjust with the changing environment and continue to bolster the ability to acquire the best vehicle given the current conditions.